



# OpenData : au-delà des publications, le partage des données de la recherche en sciences humaines. Pour qui ? Comment ?

Thierry Chanier

## ► To cite this version:

Thierry Chanier. OpenData : au-delà des publications, le partage des données de la recherche en sciences humaines. Pour qui ? Comment ?. Colloque "Open Data : enjeux communicationnels et sociétaux", Nov 2013, Clermont-ferrand, France. <edutice-00905081>

**HAL Id: edutice-00905081**

**<https://edutice.archives-ouvertes.fr/edutice-00905081>**

Submitted on 15 Nov 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# OpenData : au-delà des publications, le partage des données de la recherche en sciences humaines. Pour qui ? Comment ?

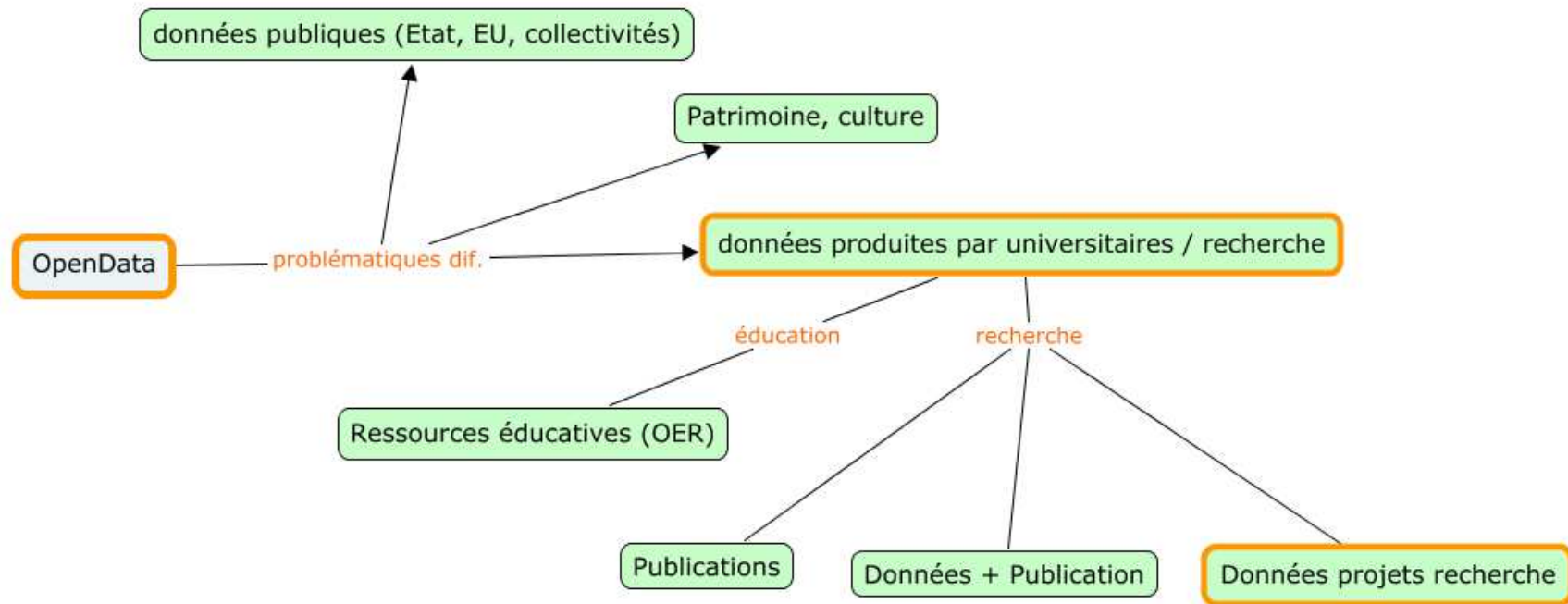
Thierry Chanier, Université Blaise Pascal



Diaporama téléchargeable à <http://goo.gl/GUEO07>

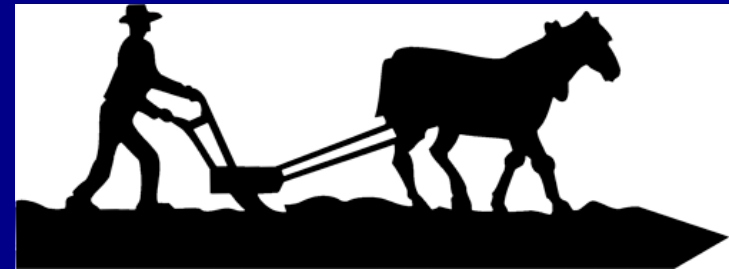
Colloque « Open Data : enjeux communicationnels et sociétaux », 7-8  
nov 2013, Clermont-Ferrand

# OpenData : différencier les enjeux



# Présuppositions

- On s'intéresse ici aux universitaires / chercheurs en tant que **producteurs de données** (et non travaillant sur des données qu'ils n'ont pas produites)
- *Open access* = accès **libre**, permanent, irrévocable sur la Toile et accès **gratuit**





Open Educational Resources (OER)

# **RESSOURCES ÉDUCATIVES LIBRES**

# Partager les ressources éducatives pour les enseignants : années 2003

- Développement des TICE sur Internet : réseaux d'enseignants, partage
- Mouvement international
- Formation enseignant Introduction C2i enseignant
- Focalisation sur métadonnées LOM (*Learning Object Metadata*)





# Aujourd'hui : apprenant , institution, réseaux sociaux

- Développement réseaux sociaux (Web 2.0) : Apprentissage formel vs informel, apprentissage auto-guidé, tout au long de sa vie, liberté pour formateurs.
- Une autre façon de faire cours ou se former ? : MOOC (accès libre et gratuit, accès de masse)
- Questionnement sur rôles institutions
- Modèles économiques à développer (coût formations)
- Repères théoriques encore flous (Know, 2013) : pour quels apprenants (autonomie demande compétences) ? Pédagogie ou seulement tourne-pages ? Institutions non nécessaires ? Mais toutes cherchent la première place – certification- , ...

MOOC: lancement de la plate-forme nationale, ça va être FUN



Coup de théâtre dans le milieu académique français. Mme Fioraso, Ministre de l'enseignement supérieur et de la recherche, annonçait mercredi matin le lancement de la plate-forme

Début dans les années 90, encore du chemin à parcourir ...

## **ACCÈS LIBRE AUX PUBLICATIONS**



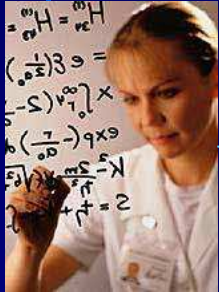
# Demandes sociétales : résultats de la recherche = un bien public

- Position des directions de recherche internationales
  - Les bénéfices que l'on tire de la recherche dépendent de l'accès aux résultats
  - Si l'accès aux recherches est restreint (aux chercheurs, citoyens, entreprises), c'est la société qui en pâtit (conséquences néfastes varient suivant type acteurs)

Une évidence ? Où est le problème ?

# Coûts recherche et publication résultats (SHS, niveau établissement)

**Salaire chercheurs : 40 K€**  
(1/2temps, coût UBP)



**Contrib équipe et données**  
antérieures : non comptées



**Relecteurs: 2\*6h**  
**500 €**



**500 à 1000 €**

**Dif revue**

**Part projet : 15 K€**  
(fonct, CDD)



**Infrastructure : 10 K€ (?)**



**Recherche**

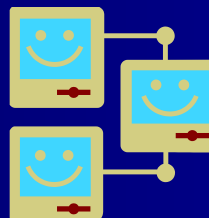
**Versions article**

**données**

**Dif OA**

**15€**

**Contrib réseau communauté :**  
non compté



# Le coût de la publication scientifique à un niveau national

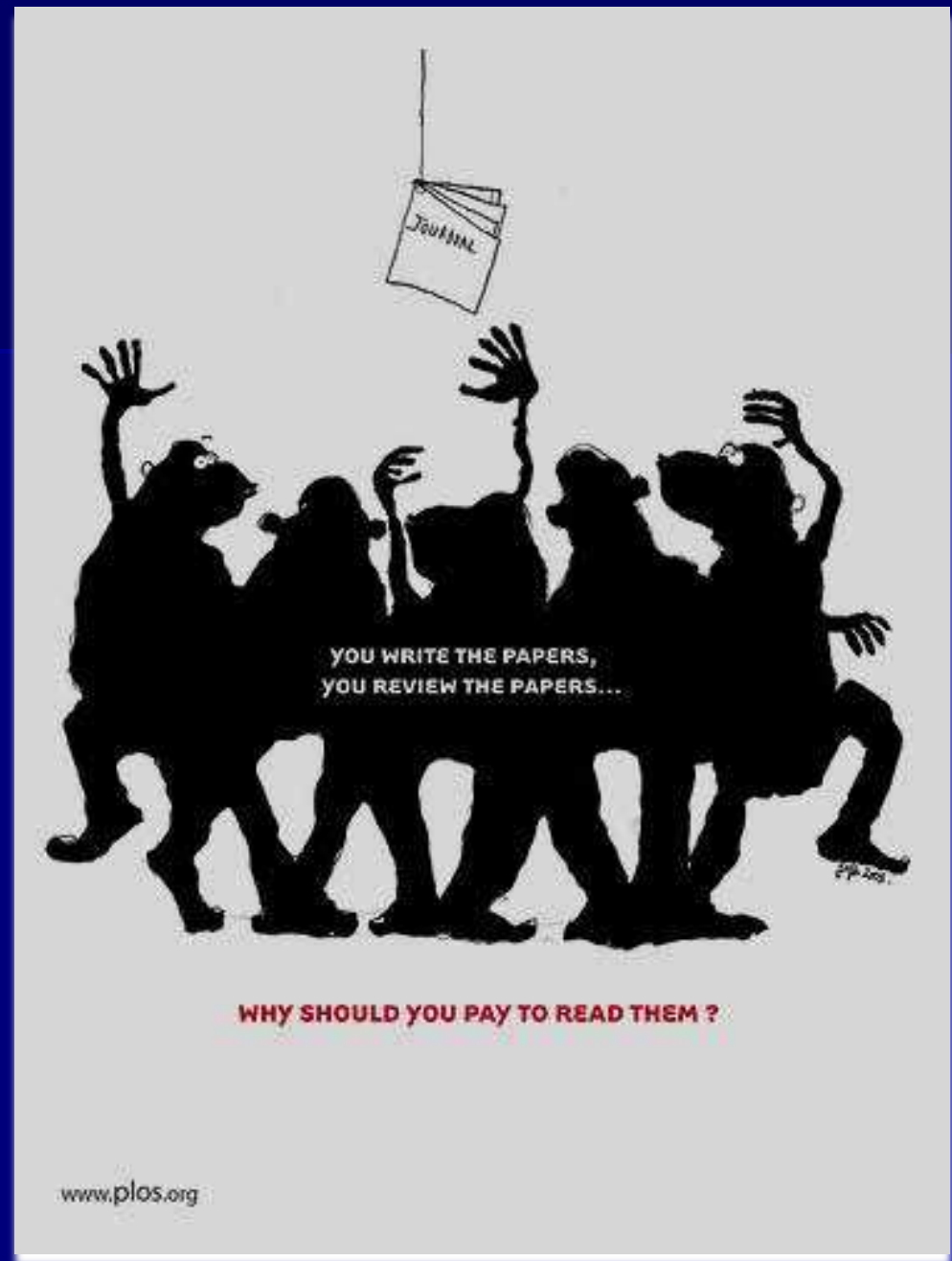
Lecture	écriture	Relecture (par paires)	Activités éditoriales	Comité éditorial	Prép prog financements (Dir. Recherche)	Sélection prog financements (Dir.)	Coûts publishers
51%	30%	3,7%	1,15%	0,13%	2,3%	0,36%	10,8%

- Études sur Danemark, Hollande, Grande-Bretagne, ici GB avec total à 8 Milliards d'euros (Houghton, 2009)
- On notera que le chercheur se trouve impliqué dans 7 des 8 activités
- Le budget total fait partie du budget total recherche et est financé sur fonds publics

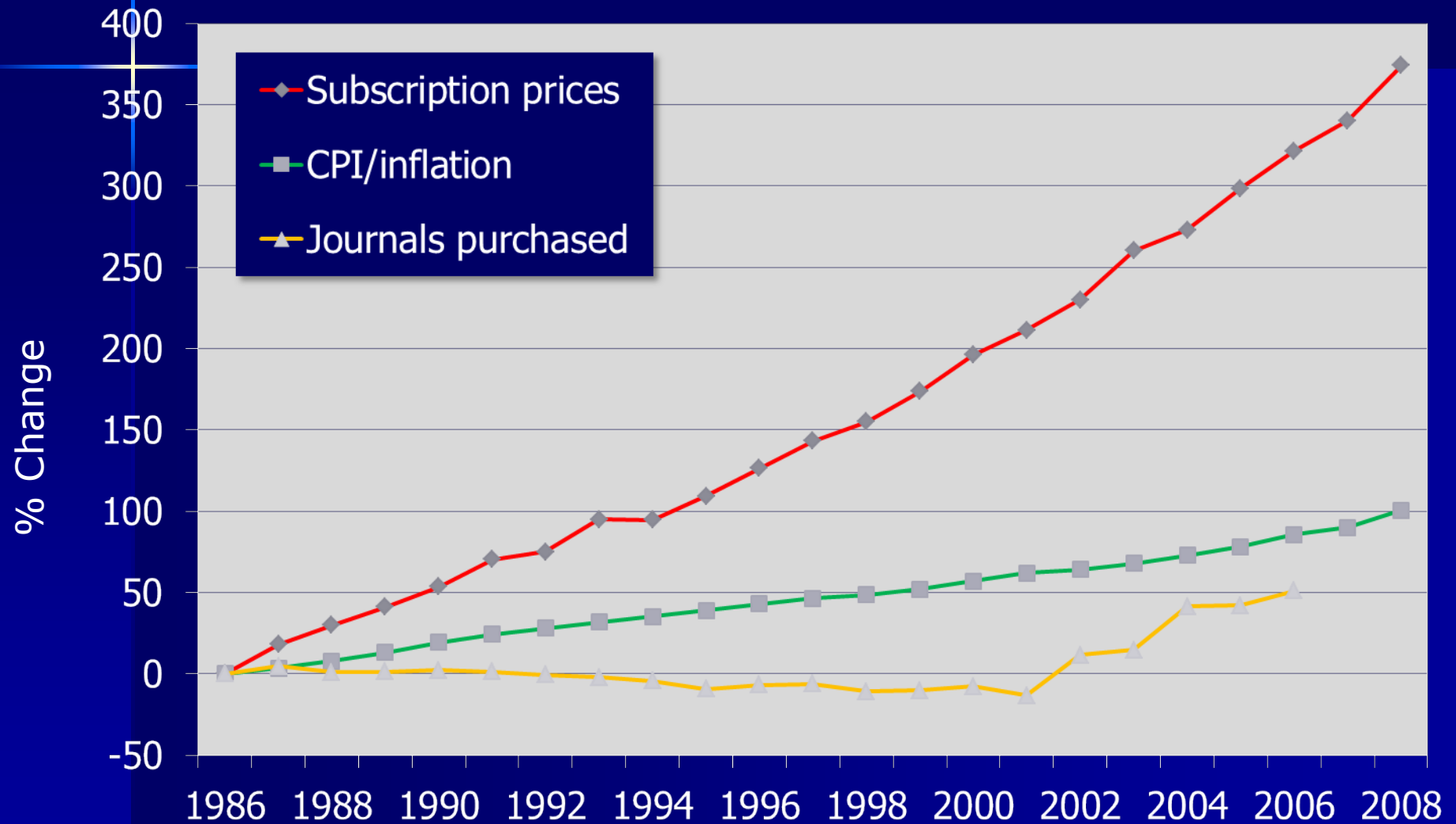
# Deux problèmes essentiels

- L'accès aux publications pour les chercheurs (et le citoyen)
- Les contrats de cession des droits d'auteur

# L'accès aux publications



# Crise des abonnements aux revues



# Grandes voies pour accès libre (*open access*) et les modèles économiques

## Voie verte (*green road*) Archives ouvertes

Archives  
institutionnelles

Archives  
thématiques

Archive  
nationale/inter.



Ici publi aussi

## Voie dorée (*gold road*) Revues AL

Auteur-payeur

Communautaire



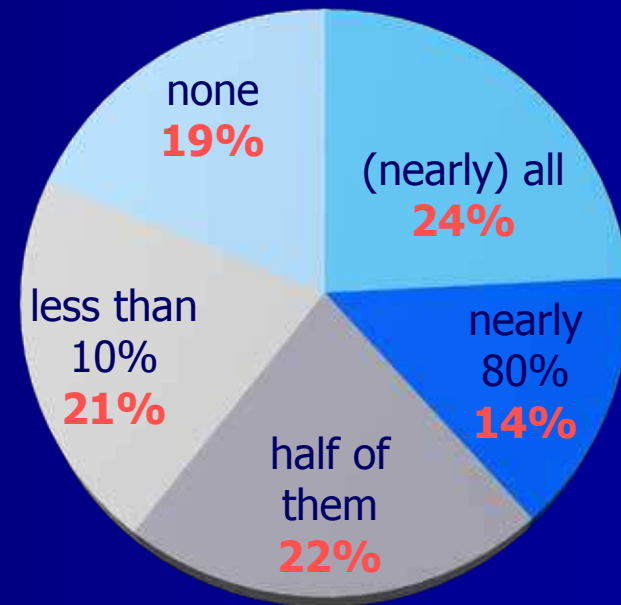


# Lente progression de l'accès libre: illustration

Milieu international  
chercheurs en TICE et langues (CALL)

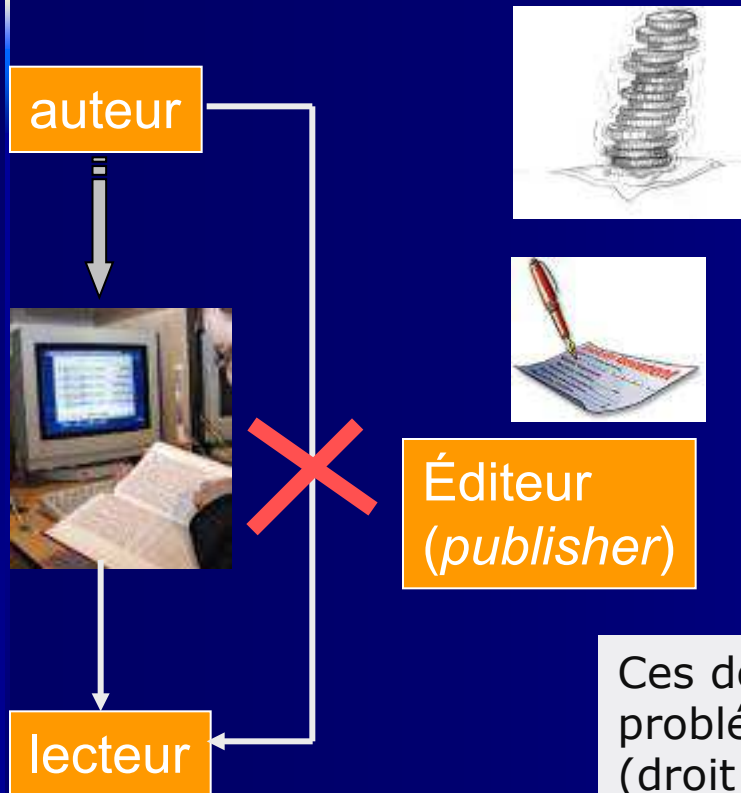
60% of researchers give open access to, at least, half of their publications.  
38% of respondents to almost all of them

## OA deposit of publications



Source : Chanier, T. (2013). *Results of the survey "CALL within the Digital Humanities: considering CALL journals, research data and the sharing of research results"* [document with associated data]. <http://edutice.archives-ouvertes.fr/edutice-00876430>

# La question des droits quand le « publisher » rémunère l'auteur



Domaines tels que :

- Les livres, contenus multimédia (musique, vidéo, etc.) hors domaines académiques,
- L'édition pédagogique,
- Les ouvrages de vulgarisation scientifique

Ces domaines ont une base et des problématiques juridiques spécifiques (droit d'auteur, copyright, licence, contrat cession, loi Adopi, etc.)

i.e. Les notions de droit doivent être abordés différemment en ce qui concerne la publication scientifique

# Contrat de cession des droits établis par publishers commerciaux

- Souvent encore abusifs (auteurs perd tout, circulation et réutilisation freinées)
- Positions qui se rapprochent de l'intimidation, pouvant aller jusqu'à l'illégalité assumée

# Un exemple récent : Taylor & Francis (2013)

- “Subject to the provisions of the accompanying publishing agreement information form, **you assign to us with full title guarantee all rights of copyright and related rights in your Article**. So that there is no doubt, this assignment includes the assignment of the rights (i) to publish, reproduce, distribute, display and store the Article worldwide **in all forms, formats and media now known or as developed in the future**, including print, electronic and digital forms, [...]
- You are permitted to use the material in the ways described in the Schedule of Author's Rights [...] **These are rights which are personal to you and cannot be transferred by you to anyone else »**

# Soyons optimistes !

- “The Commission proposes to make open access to scientific publications a general principle of Horizon 2020, building on the already existing activities in FP7 (e.g. eligibility of open access publishing costs, embargo for 'Green' open access of six to twelve months).
  - there should be open access to publications resulting from publicly funded research as soon as possible, preferably immediately and in any case no later than six months after the date of publication, and twelve months for social sciences and humanities;
  - licensing systems contribute to open access to scientific publications resulting from publicly-funded research in a balanced way, in accordance with and without prejudice to the applicable copyright legislation, and encourage researchers to retain their copyright while granting licences to publishers;

[http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/background-paper-open-access-october-2012\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/background-paper-open-access-october-2012_en.pdf)

# **Soumettre et publier articles et données**

Projet Datapublication.org

# Situation aujourd'hui

- Articles reposant sur des données de recherche publiés sans les données
- Imaginez une enquête sans que l'on puisse voir le questionnaire, sa formulation, les données recueillies, les calculs : peut-on faire confiance ? Puis-je comparer ces résultats à d'autres ?
- Comment les relecteurs peuvent-ils faire un travail de certification de la qualité scientifique ?



# Objectifs

- Relecteurs accèdent aux données déposées par l'auteur pendant processus de relecture
- Lorsque l'article est accepté, les données sont publiées
- Le lecteur (chercheur) peut accéder à ces données afin de répliquer, joindre ces données aux siennes pour étendre étude
- Auteur gagnant ! Deux références pour son travail : données auront une référence distinctes (mais reliées à l'article)

# Début en Tice en France : projet Datapublication .org

[http://sticef.univ-lemans.fr/num/vol2012/05-guichon/sticef\\_2012\\_guichon\\_05.htm](http://sticef.univ-lemans.fr/num/vol2012/05-guichon/sticef_2012_guichon_05.htm)

<http://datapublication.tge-adonis.fr/data/d-001-102>

## “Data publication - publish articles and research data -

### Données de l'enquête sur les usages des TIC par les lycéens.

#### Information

Year: 2012

Author(s): Nicolas Guichon

Provide by: **Nicolas Guichon**

Keywords: usages numériques, fracture numérique, compétences, Technologies de l'Information et de la Communication Technologies (ICT), ICT usage, digital divide, competences

#### Description

Un questionnaire concernant l'origine (profession des parents, lycée, équipement informatique, B2I), l'usage perso administré au printemps 2011 à 1300 élèves de première en France métropolitaine. L'enquête avait pour principal but lycéens de l'enseignement secondaire général en France. Elle a concerné 10 lycées issus de diverses zones géographi périphérie urbaine ou centre d'une grande ville.

#### Restrictions

Les données issues de ce questionnaire sont destinées à la communauté scientifique. Toute utilisation suppose d diffusion doit faire l'objet d'une demande particulière à l'auteur. L'utilisation commerciale n'est pas autorisée.

#### Data file

To download the files click on their name. The downloadable files are the formats provided by the author. The file in a request must be do from the author.

<a href="#">Questionnaire_FINAL-Guichon.pdf</a> application/pdf - 25.4 ko	Other	document (f questionnai
<a href="#">Enquête Guichon2011.xls</a> application/vnd.ms-excel - 1.23 Mo	Other	fichier texte ré
<a href="#">Codage1.csv</a> application/vnd.ms-excel - 3.74 ko	Data	
<a href="#">Codage2.csv</a> application/vnd.ms-excel - 6.07 ko	Other	
<a href="#">Etablissement.csv</a>		

technologies  
on et  
ication pour  
t la Formation

[version pleine page](#)

[version à télécharger \(pdf\)](#)

Volume 19, 2012  
Article de recherche

### Les usages des TIC par les lycéens - déconnexion entre usages personnels et usages scolaires

[Nicolas GUICHON \(ICAR, Lyon 2\)](#)

■ **RÉSUMÉ** : Cette recherche, qui adopte la perspective de la sociologie des usages, s'appuie sur une enquête à la fois par questionnaire et par entretiens pour sonder les usages numériques des lycéens de l'enseignement secondaire général en France. Deux objectifs sont visés : d'une part, grâce aux données empiriques obtenues, un état des lieux des usages numériques des jeunes est conduit. D'autre part, cette étude investigate de quelles façons les outils numériques sont utilisés pour le travail à la maison et pour l'apprentissage des langues étrangères. Les résultats mettent au jour une déconnexion entre usages des Technologies de l'Information et de la Communication (TIC) entre la sphère privée et la sphère scolaire.

■ **MOTS CLÉS** : usages numériques, fracture numérique, compétences, Technologies de l'Information et de la Communication (TIC)

■ **DONNÉES ASSOCIÉES** : Accessibles via le lien permanent

([Datapublication.org](#) : [Tge-adonis.fr](#).)

<http://hdl.datapublication.org/11107/d-001-102>

# Ce que nous dit l'Europe

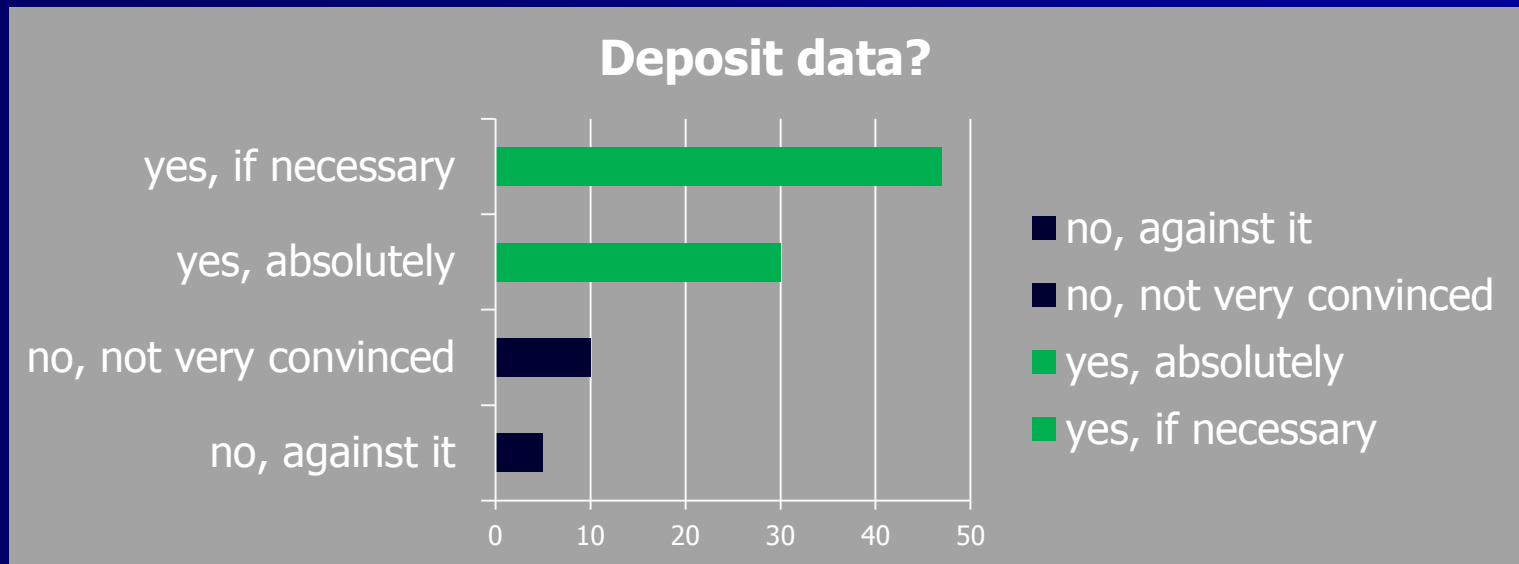
## 8. What does the Commission propose to do about open access to data in Horizon 2020?

The Commission proposed in its Communication 'Towards better access to scientific information' to develop a pilot on open access to data, primarily those data underlying (open access) scientific publications. The areas covered by the pilot should be discussed together with the thematic Units within the Commission and their stakeholders.

COMMISSION RECOMMENDATION of 17.7.2012 on access to and preservation of scientific information : [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/recommendation-access-and-preservation-scientific-information\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf)

# Qu'en pensent les chercheurs ?

**Suppose you submit a paper to a journal which is based on research data, would you be willing to deposit your data for the review process?**



83% of survey participants would accept to deposit their data in order to be taken into account for this data to be considered during the review process.

# Un projet à l'étude au niveau international en SHS

- Plusieurs revues internationales en Tice et langues
- Et quelques autres revues en SHS
- Poser projet européen, type Datapublication.org avec soutien DARIAH (Humanités numériques)



# **OPENDATA SUR LES DONNÉES DE LA RECHERCHE**



# Des habitudes tenaces en sciences humaines

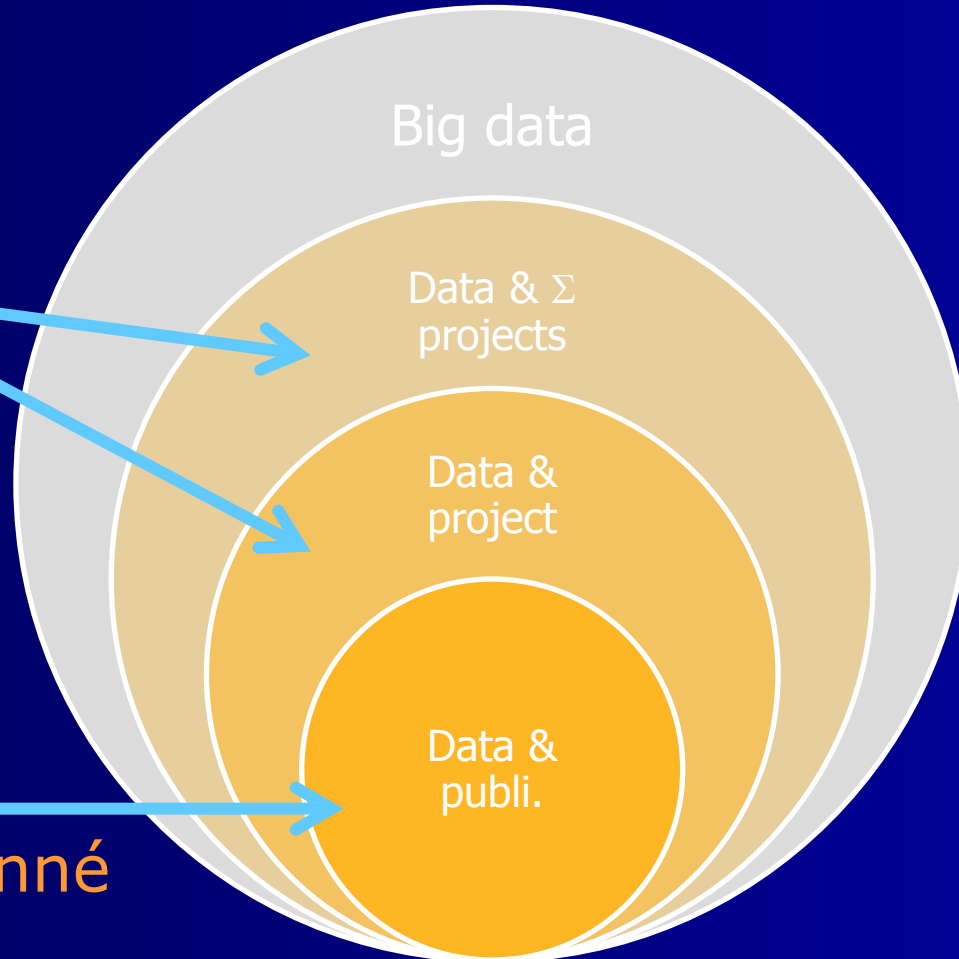
- Je publie et jette mes données ; argent public gaspillé
- Je ferme l'accès à mes analyses et jette la clef
- Je garde mes vieilles données dans les placards
  - «vous voulez des données, elles sont là !»
  - ... mais elles sont inutilisables et perdues à jamais





# Differentes couvertures pour les données

En linguistique, on les baptise « corpus »



Déjà mentionné

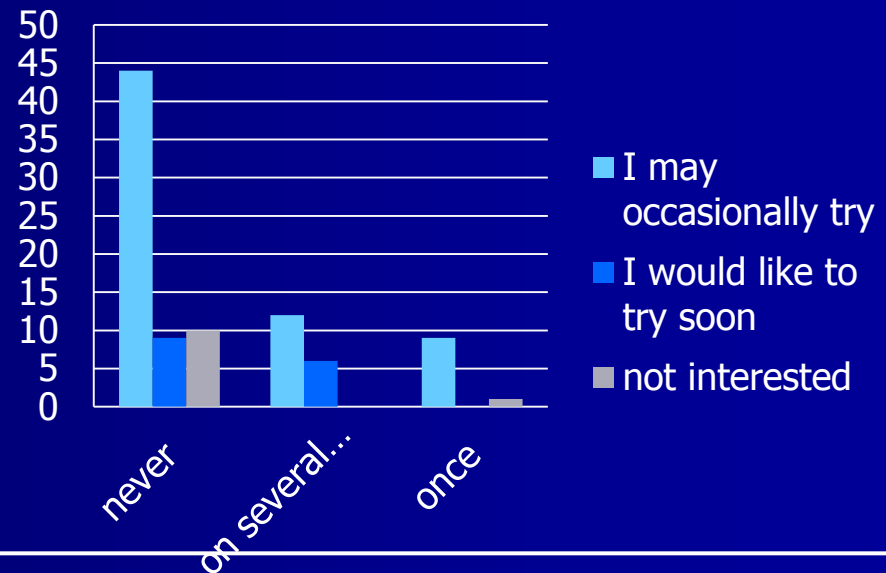
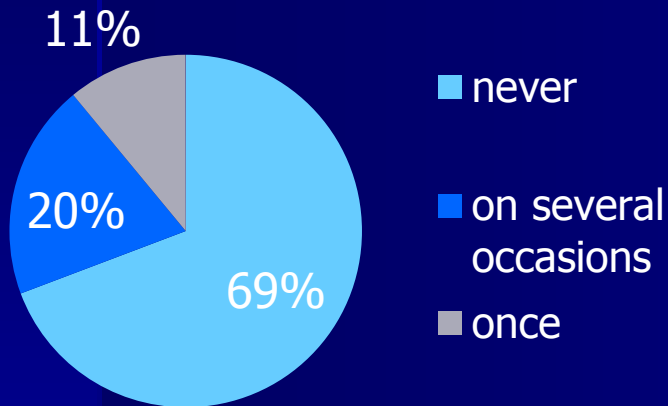
# Opendata def

- "Open data is data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike." OpenDefinition.org

# Qu'en pensent les chercheurs ?

If research data from learning situations was available in open-access formats, would you be interested in doing research on this data that you did not personally collect?

**Use of other's data**



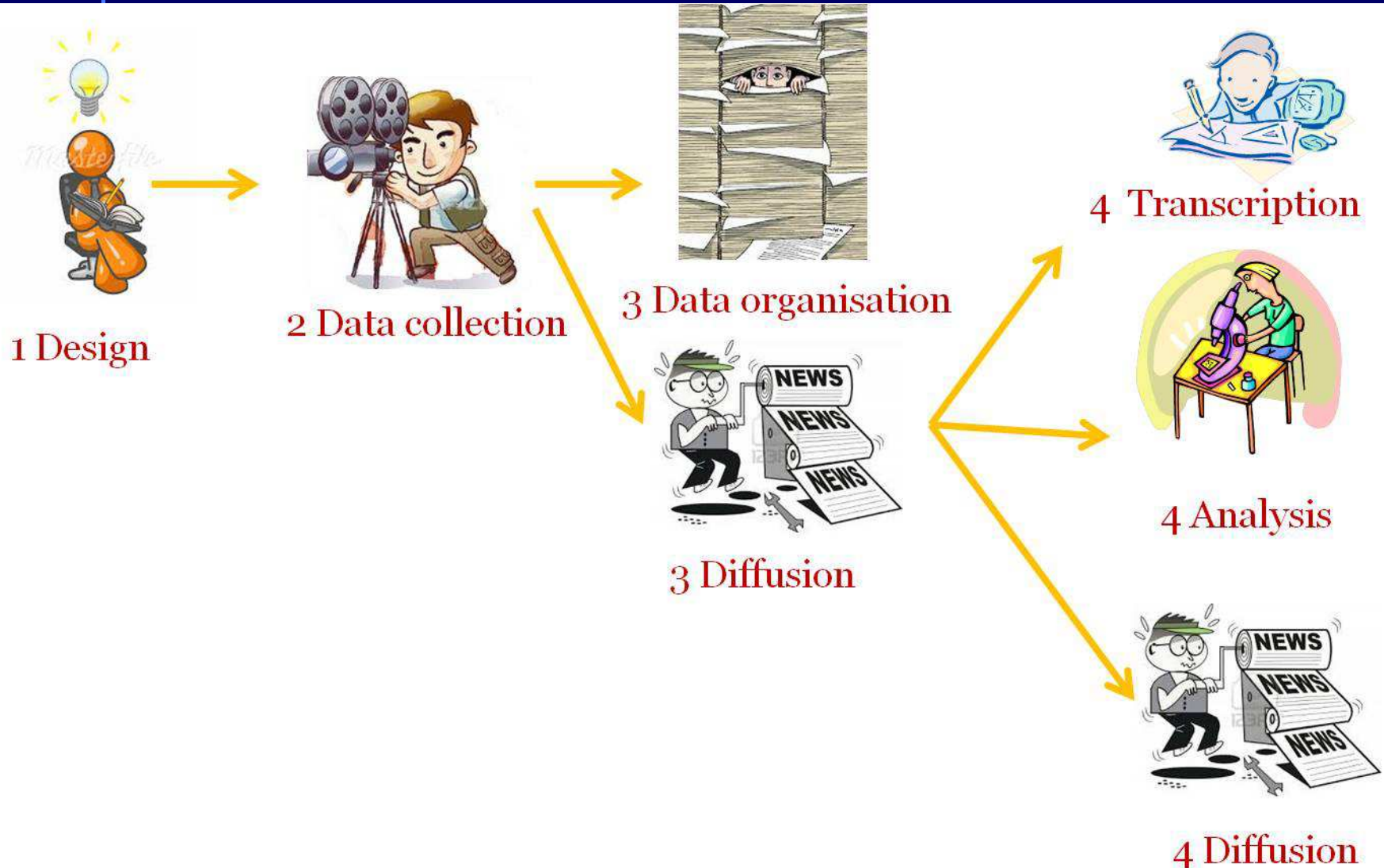
Source : Chanier, T. (2013).

88% of researchers find that doing some research on learning situations they did not design nor tutor may be of interest (in blue),

# Partager ce n'est pas seulement vouloir

- Cela implique une mutation dans les façons de travailler
- Organiser, structurer ses données, les renseigner, le tout de façon standard
- Développer des outils d'analyse interoperables
- Régler les questions d'éthique et de droits

# Former les chercheurs sur l'ensemble du flux de production



# Un exemple : banque de corpus Mulce diffusée à la MSH Clermont par LRL



Open Data:  
<http://opendefinition.org/guide/>



For usage:  
licence



For participants:  
Informed  
consent form  
+  
Anonymization  
process

Note : Incoherence between licences on our sites: changes are not yet achieved

UK-FR



eng

UK-FR-USA



eng, fre

DE-FR



deu, fra

USA-FR



fra

UK-USA-SG



math

FR



fra

FR



eng, fra

UK-FR



fre

CO-FR



fre, spa





# CoMeRe (Communication Médiée par les Réseaux): a reference corpus of French CMC (2013-14)



<http://corpusecrits.corpus-ir.fr/>

Project supported by the national consortium *Corpus-écrits*, sub-part of *Huma-Num*, and *Ortolang* (French correspondent to DARIAH)

Documentation and activities : <http://comere.org>

**People:** 14 research. from 8 research units. Coord: Chanier, T (Clermont), Poudat, C. & Sagot, B (Paris), Longhi, J. (Cergy), Antoniadis, G. (Grenoble)

**Objective:** Kernel corpus assembling existing corpora of different CMC genres and new corpora build on data extracted from the Internet. These heterogeneous corpora will be structured and processed in a uniform way, complemented with metadata. CoMeRe will be released as OpenData through the national infrastructure Ortolang, following constraints which will be reused for the forthcoming “*Corpus de Référence du Français*”.

**Passage à l'échelon national**

# Passage à l'échelon européen (2013-2014)

- Organiser ses données de façon standard



- groupe européen (Italie, Allemagne, Hollande, France) TEI-CMC pour étendre TEI aux communications médiées sur réseaux
- Soutien de DARIAH



# Critères pour être OpenData

- **"Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Reuse and Redistribution:** the data must be provided under terms that permit **reuse and redistribution including the intermixing with other datasets**. The data must be machine-readable.
- **Universal Participation:** everyone must be able to use, reuse and redistribute – there should be no discrimination against fields of endeavor or against persons or groups. For example, **'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.** "OpenDefinition.org

# **LA QUESTION DES LICENCES D'UTILISATION**

# L'accès libre n'est pas automatique sur Internet

- I seriously doubt that we would create the Web today—at least if policy makers and market incumbents understood what the technology might become early enough to stop it. (p.278)
- Almost everything on the Internet is copyrighted, even if its creators do not know that and would prefer it to be in the public domain. (p. 26)

(Boyle, J.2008, *The Public Domain: Enclosing the Commons of the Mind*) Boyle is one of the creators of the Creative Common – CC project

# Pourquoi doit-on se soucier des licences ?

- "In most jurisdictions there are intellectual property rights in data that prevent third-parties from using, reusing and redistributing data without explicit permission. Even in places where the existence of rights is uncertain, it is important to apply a license simply for the sake of clarity. Thus, **if you are planning to make your data available you should put a license on it** — and if you want your data to be open this is even more important."  
OpenDefinition.org

# On croit que c'est libre, mais cela ne l'est pas : exemple ELFA

LICENSE AGREEMENT

<https://elomake.helsinki.fi/lomakkeet/43518/lomake.html>

## CLARIN RESTRICTED END-USER LICENCE (RES)

**Copyright holder:** Anna Mauranen; The Faculty of Arts, PO Box 3, 00014 University of Helsinki; anna.mauranen@helsinki.fi

**Resource:** ELFA – English as a Lingua Franca in Academic Settings

The Copyright holder grants the End-User a free, non-exclusive and perpetual (for the duration of the copyright) right to **use and make copies of the Resource for personal use** as such, as modified, or as part of a combined work. The permission applies to all known modes and means of communication and includes a right to make modifications enabling the use of the Resource on other devices and in other formats. The purpose of the Resource use must be outlined in the research plan. Distribution of copies is not allowed.

The Resource may be used for publishing articles and analyses in scientific journals, conference presentations or other similar forums. In such publications, proper reference to the resource should be given. [INF: If the Resource is used as material for a scientific publication, the Copyright holder is to be informed.]

[PD: Due to the nature of the material, the Resource should be handled with care in order to respect the privacy of the personal data. If samples of the data are published, they must be anonymized according to best practices.]

[NC: The Resource may not be used for gaining economic benefit. Government-funded or non-profit research projects, e.g. projects funded by Academy of Finland or TEKES, are not regarded as gaining economic benefit even if a portion of the financing is contributed by companies.]

[ReD: A revised or enhanced version of the Resource may be redeposited with the CLARIN Service. The conditions for redepositing will be agreed separately.]

- Open access, but for personal use (hence not for research)
- Important restriction (NC), where are the sound files?



# Les licences Creative Commons - CC (pas toutes compatibles avec l'OpenData)



## Attribution CC BY

Cette licence permet aux autres de distribuer, remixer, arranger, et adapter votre œuvre, même à des fins commerciales, tant qu'on vous accorde le mérite de la création originale en citant votre nom. C'est le contrat le plus souple proposé. Recommandé pour la diffusion et l'utilisation maximales d'œuvres licenciées sous CC.

[Voir le Résumé Explicatif](#) | [Voir le Code Juridique](#)



## Attribution - Partage dans les Mêmes Conditions CC BY-SA

Cette licence permet aux autres de remixer, arranger, et adapter votre œuvre, même à des fins commerciales, tant qu'on vous accorde le mérite en citant votre nom et qu'on diffuse les nouvelles créations selon des conditions identiques. Cette licence est souvent comparée aux licences de logiciels libres, "open source" ou "copyleft". Toutes les nouvelles œuvres basées sur les vôtres auront la même licence, et toute œuvre dérivée pourra être utilisée même à des fins commerciales. C'est la licence utilisée par Wikipédia ; elle est recommandée pour des œuvres qui pourraient bénéficier de l'incorporation de contenu depuis Wikipédia et d'autres projets sous licence similaire.

[Voir le Résumé Explicatif](#) | [Voir le Code Juridique](#)



## Attribution - Pas de Modification CC BY-ND

Cette licence autorise la redistribution, à des fins commerciales ou non, tant que l'œuvre est diffusée sans modification et dans son intégralité, avec attribution et citation de votre nom.

[Voir le Résumé Explicatif](#) | [Voir le Code Juridique](#)



## Attribution - Pas d'Utilisation Commerciale CC BY-NC

Cette licence permet aux autres de remixer, arranger, et adapter votre œuvre à des fins non commerciales et, bien que les nouvelles œuvres doivent vous créditer en citant votre nom et ne pas constituer une utilisation commerciale, elles n'ont pas à être diffusées selon les mêmes conditions.

[Voir le Résumé Explicatif](#) | [Voir le Code Juridique](#)



## Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions CC BY-NC-SA

Cette licence permet aux autres de remixer, arranger, et adapter votre œuvre à des fins non commerciales tant qu'on vous crédite en citant votre nom et que les nouvelles œuvres sont diffusées selon les mêmes conditions.

[Voir le Résumé Explicatif](#) | [Voir le Code Juridique](#)





## Attribution - Pas d'Utilisation Commerciale - Pas de Modification CC BY-NC-ND

Cette licence est la plus restrictive de nos six licences principales ; elle autorisant les autres qu'à télécharger vos œuvres et à les partager tant qu'on vous crédite en citant votre nom, mais on ne peut les modifier de quelque façon que ce soit ni les utiliser à des fins commerciales.

# 2 licences sur les données compatibles OpenData

## Conformant Data Licenses



License	Domain	By	SA	Comments
Open Data Commons Public Domain Dedication and Licence (PDDL)	Data	N	N	Dedicate to the Public Domain (all rights waived)
Open Data Commons Attribution License	Data	Y	N	Attribution for data(bases)
Open Data Commons Open Database License (ODbL)	Data	Y	Y	Attribution-ShareAlike for data(bases)
Creative Commons CCZero	Content, Data	N	N	Dedicate to the Public Domain (all rights waived)

# CC0 and PDDL

- CC0 : en tant que créateurs, je peux avoir des droits (droits sur les modèles ou les données) As creators, I may have had some rights (rights on models, rights on data, etc.) , je les abandonne (de façon permanent , et irrévocable)
- PDDL : je ne mentionne même pas que je pourrai avoir des droits

## About CC0 — “No Rights Reserved”



CC0 enables scientists, educators, artists and other creators and owners protected content to waive those interests in their works and thereby place possible in the public domain, so that others may freely build upon, enhance and reuse the

### You are free:



To Share: To copy, distribute and use the database.



To Create: To produce works from the database.



To Adapt: To modify, transform and build upon the database.

### As long as you:



Blank: This section is intentionally left blank. The PDDL imposes your use of the PDDL licensed database.

# Question angoissante pour les chercheurs



- Que va-t-il se passer si je n'indique plus mes droits d'auteur (copyright) sur les données (on ne parle pas ici des publications) ?
- Va-ton me voler mon travail ? Ne va-t-on plus me citer ?

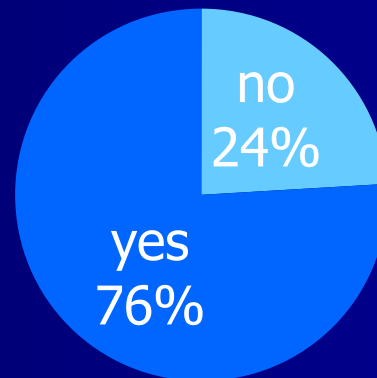
# Ne pas confondre attribution droits d'auteur (IPR) avec façons de citer et référer

- Contestation sur droits d'auteur se règle au tribunal
- Dans le monde académique, procédures de relecture : un auteur qui utilise sans citer les travaux / données des autres voit son article rejeté.
- Développer références aux données et les rendre visibles (métadonnées, OAI-MPH)
  - 1) creator of the corpus
    - Wigham, C.R. (2013). *Distinguished Corpus: Interplay between textchat and audio modalities during the Second Life Reflective Sessions*. Mulce.org : Clermont Université. [oai : mulce.org:mce-archi21-modality-textchat ; <http://repository.mulce.org>]
  - 2) creator and editor
    - Stahl, Gerry ; Weimar, Steve ; Shumar, Wes (2009). *LETEC Corpus Virtual Math Team*. Reffay, C. (editor). Mulce.org : Clermont Université. [oai : mulce.org:mce-vmt-letec-teamc ; <http://repository.mulce.org>]

# Que pensent les chercheurs ?

Occasionally, would you be ready to waive your intellectual property rights when you deposit data (data, not publications), provided that your data can be clearly cited / referred to?

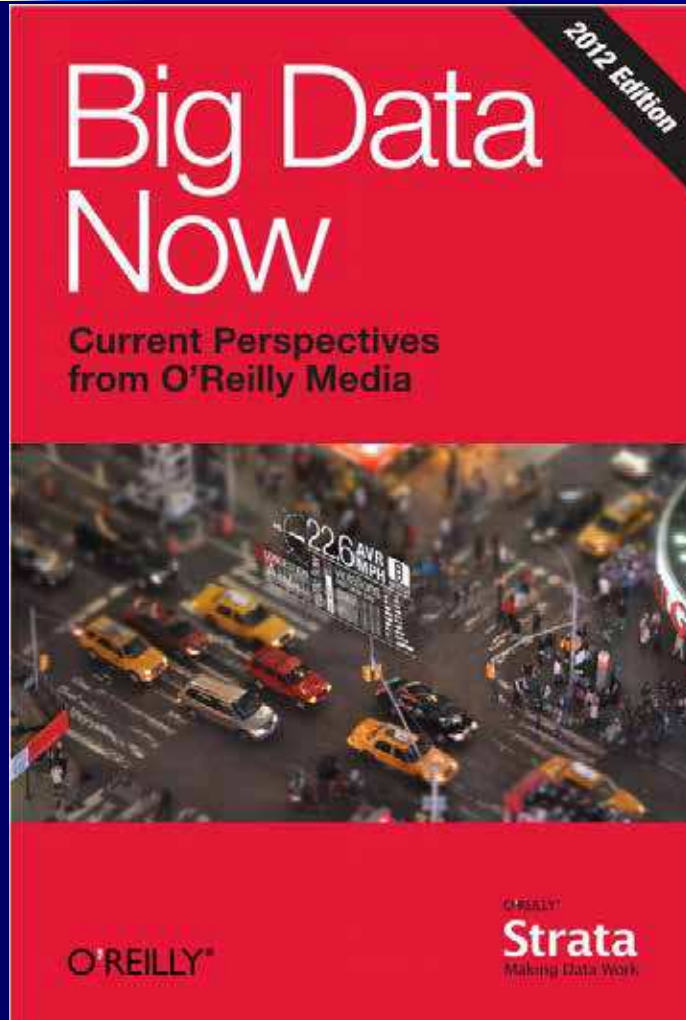
## Waive your IPR?



Source : Chanier, T. (2013).



# Qu'est ce qui se prépare après ?



O'Reilly a inventé  
récemment le terme  
« Web2.0 »

Les chercheurs en SHS  
peuvent se dire ...





## *Big data* cela veut dire ...

- Mélanger des données afin d'obtenir de grands ensembles (Tera octets)
- 0 base de données, pas de structure sur les données
- Calcul intensif (*cloud computing*, apprentissage automatique)
- Façon montante de construire les questions de recherche et les modèles
- Certains secteurs SHS commencent à bouger

# Atelier BigData en SHS

## BIG HUMANITIES

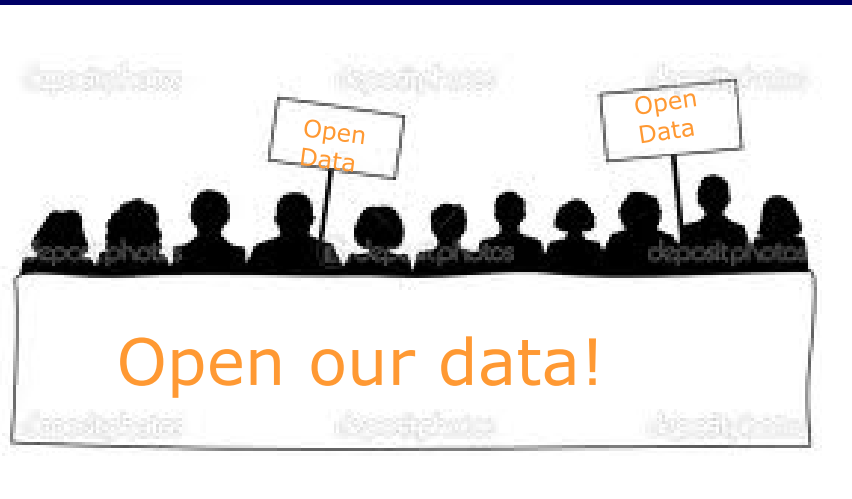
---

[Home](#) [Programme Committee](#)

---

### IEEE BIGDATA 2013: WORKSHOP ON BIG HUMANITIES, 8 OCTOBER 2013

The Workshop on Big Humanities will be held in conjunction with the [2013 IEEE International Conference on Big Data \(IEEE BigData 2013\)](#), which will take place between 6-9 October 2013 in Silicon Valley, California, USA, and which provides a leading international forum for disseminating the latest research in the growing field of "big data".



# Recommendations

## ■ Actions

- Libérer ses données (si éthique est OK – faire le nécessaire pour anonymisation)
- Choisir la licence imposant le moins de restrictions
- Citer les données des autres et les vôtres dans les publications
- Lister ces références dans vos travaux (évaluation)

## ■ Implications

- La reconnaissance viendra progressivement (de la part des institution et des collègues)
- La recherche progressera (re-analyses, couverture étendue des bases d'analyse)
- Les données seront réutilisées dans d'autres domaines que ceux des disciplines d'origine

Merci pour votre attention ! Questions ?  
Réactions ?

Diaporama téléchargeable à <http://goo.gl/GUEO07>

Thierry.chanier at univ-bpclermont.fr

<http://lrl.univ-bpclermont.fr/spip.php?rubrique98>